Examining Registration Delays in Mobile based Social Networks

Péter Ekler*, Tamás Lukovszki** and Hassan Charaf*

* Budapest University of Technology and Economics, Department of Automation and Applied Informatics, Hungary

** Faculty of Informatics, Eötvös Loránd University, Budapest, Hungary

peter.ekler@aut.bme.hu, lukovszki@inf.elte.hu, hassan.charaf@aut.bme.hu

***Abstract*— The popularity of web based social networks is rapidly increasing nowadays. These solutions allow managing relationships online. There are also several mobile solutions to these networks but they are mainly limited to reaching the services of the social network from the mobile device. The fact, that the phonebook of the mobile phones also represents social relationships, can be used for detecting additional relationships in the social network automatically, which has several benefits. For this a similarity detecting algorithm is required, which runs in the background on demand, on the server. When a member registers and synchronizes her or his entire phonebook the system should detect similarities as fast as possible. In this paper we introduce our mobile based social network. We discuss about similarity detecting based on phonebook sizes and we show a queuing theory based model regarding to registration delays.**

I. Introduction

In the last decade Internet related technologies have developed rapidly. As a reason of this growth, new types of services and applications have appeared. One of the most popular solutions are social network sites. Since their introduction, social network sites such as Facebook, Myspace and LinkedIn have attracted millions of users, many of whom have integrated these sites into their daily practices. Facebook is the second most popular visited websites on the Internet behind Google [1].

According to new statistics [2] Facebook has more than 400 million active users, 50% of the active users log on to Facebook in any given day, more than 35 million users update their status each day and an average user spends more than 55 minutes per day on Facebook. Facebook began in early 2004. Above statistics show that such popular social networks can have a huge growth which has to be considered during the design of any social network site.

Mobile phones and mobile applications are another hot topic nowadays. Facebook statistics also show that there are more than 65 million active users currently accessing Facebook through their mobile devices. People that use Facebook on their mobile devices are almost 50% more active on Facebook than non-mobile users. The increasing capabilities of mobile devices allow them to participate in social network applications as well. Mobile phone support in social networks are usually limited mainly to photo and video upload capabilities and access to the social network using the mobile web browser.

In our research we consider the fact, that the phonebook of the mobile device also describes social relationships of its owner. Discovering additional relations in social networks is beneficial for sharing personal data or other content. Given a solution that allows us to upload as well as download our contacts to and from the social networking application, we can completely keep our contacts synchronized so that we can see all of our contacts on the mobile phone as well as on the web interface. In addition to that, if the system detects that some of the private contacts in the phonebook are similar to another registered members of the social network (i.e. may identify the same person), it can discover and suggest social relationships automatically. In the rest of this paper we refer to this solution as a mobile based social network.

Discovering and handling such similarities in mobile based social networks needs additional computational resources but it has a lot of benefits. With the help of detected similarities the system can keep the phonebooks always up-to-date with information provided by the other people. In this paper we briefly describe mobile based social networks, we show how important it is to prepare for similarity handling during registration and we propose a model for registration delay where we were able to apply Kleinrock's queuing model [14].

The described mobile based social network architecture is actually applied in the Phonebookmark project at Nokia Siemens Networks. We took part in the implementation and before public introduction it was available for a group of general users from April to December of 2008. It had 420 registered members with more than 72000 private contacts, which is a suitable number for analyzing the behavior of the network. During this period we have collected and measured different types of data related to the social network and its behavior.

The rest of the paper is organized as follows. Section 2 describes related work in the field of social networks and mobile solutions. Section 3 briefly summarizes the architecture of mobile based social networks. Section 4 describes similarity detecting considering phonebook sizes. Section 5 shows a model with measurements for estimating registration delays related to similarity detecting. Finally Section 6 concludes the paper and proposes further research areas.

II. Related work

In [3] the authors have defined social network sites as web - based services that allow individuals to (1) construct a public or semi - public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system. The nature and nomenclature of these connections may vary from site to site.

According to this definition, the first recognizable social network site launched in 1997. SixDegrees.com allowed users to create profiles, list their Friends and, from beginning of 1998, surf the friends lists. Later, social networks have developed rapidly and the number of features have increased. Nowadays most sites support the maintenance of preexisting social networks, others help strangers to connect to people based on shared interests or activities. The sites have different ways to incorporate new information and communication tools, such as mobile connectivity, blogging, and photo/video sharing.

As the functionality of the social network sites has been grown, the number of users have increased rapidly. Handling the extending number of users efficiently is a key issue as it was visible in case of Friendster. Friendster was launched in 2002 as a social complement to Ryze. As Friendster's popularity surged, the site encountered technical and social difficulties. Friendster's servers and databases were ill−equipped to handle its rapid growth, and the site faltered regularly, frustrating users who replaced email with Friendster.

Newman et al. [4] describe some novel uniquely solvable models of the structure of social networks based on random graphs with arbitrary degree distributions. They give models both for simple unipartite networks such as acquaintance networks and bipartite networks such as affiliation networks. They compare the predictions of their models to data from a number of real-world social networks and find that in some cases, the models show high correlation with the data, whereas in others the correlation is lower, perhaps indicating the presence of additional social structure in the network that is not captured by the random graph.

Mislove et al. [5] studied the graph properties of several online real-world social networks. Their paper presents a large-scale measurement study and analysis of the structure of multiple online social networks. They examined data gathered from four popular online social networks: Flickr, YouTube, LiveJournal, and Orkut. They crawled the publicly accessible user links on each site, obtaining a large portion of each social network's graph. Their data set contains over 11.3 million users and 328 million links. Their measurements show high link symmetry, i.e. users tend to receive as many links as they give. The distribution of the node degree in the observed networks follows a power-law with high symmetry.

In [6], the graph structure of the Web has been investigated and it has been shown that the distribution of in- and out-degrees of the Web graph are well approximated by power law distributions. Nazir et al. [7] showed that the in- and out-degree distribution of the interaction graph of the studied Myspace applications also follow such distributions.

Kumar et al. [8] presented measurements related to online social networks. They analyzed data of five million people and ten million friendship links, annotated with metadata capturing the time of every event in the life of the network. Their measurements expose a surprising segmentation of these networks into three regions: singletons who do not participate in the network; isolated communities which overwhelmingly display star structure; and a giant component anchored by a well-connected core region which persists even in the absence of stars. They also presented a simple model of network growth which captures these aspects of component structure.

During the development period of Phonebookmark, we have observed other phonebook related solutions on the web. Zyb [9] and Plaxo [10] allow for synchronizing with mobile phones and managing the contacts using a web browser. Xing [11] has also mobile access, but focuses more on business relationships. There is no automatic similarity detection in these systems, thus there is no notification like in Phonebookmark when one of a user's phonebook contacts becomes (or already is) a member of the system.

Analyzing the structure of a social network is a key issue if we consider a large number of users. The key difference between previous research and our work that we bring together the online and "mobile" relationships based on the phonebook of the mobile devices.

III. ANALYZING MOBILE BASED SOCIAL NETWORKS

Mobile based social networks are extending the well-known social network sites. They have a similar web user interface, but they add several major mobile phone related functions to the system. In [12] we have defined the structure of such networks in detail. In this section we just give a brief introduction as it is necessary to understand the rest of this paper.

In the following we consider social networks as graphs. In case of general social networks, nodes represent registered members and edges between them represent social relationships (e.g. friendship). After this we should notice that each member has a private mobile phone with a phonebook (Figure 1).

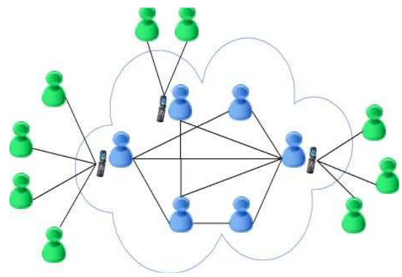

Figure 1. Phonebook-enabled social network

Phonebook contacts are represented by a new type of nodes in the graph (see Figure 1, contacts linked to the phone). The edges between private phonebook contacts and members represent which member "owns" those private contacts.

One of the key advantages of mobile based social networks is that they allow real synchronization between private phonebook contacts and the social network. In order to enable such mechanism a similarity detecting algorithm is required. This algorithm compares pairs of person entries (members and private contacts too) and determines a weight for a pair, which express how likely the two entries represent the same person. The details of the algorithm are described in [13].

Figure 2 represents the graph structure if the similarity detecting algorithm has finished comparing the relevant person entries. The dotted edges between member and private contacts represent detected similarities and broken lines between two private contacts illustrate possible duplications in the phonebooks. Duplications are detected as a positive side effect of the similarity detecting algorithm.

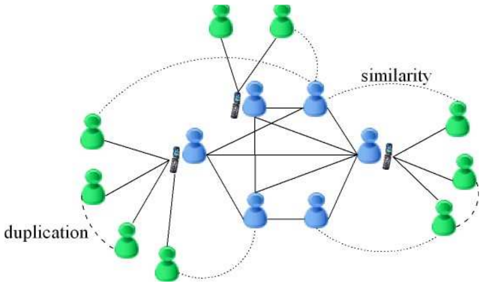

Figure 2. Detected similarities and duplications

After similarities and duplications are detected there is a semi-automatic step, the members who have private contacts in their phonebook, which are detected as similar to other members - have to decide whether the detected similarities are relevant ones. In addition to that, members can also decide about the relevancy of detected duplications in their phonebooks. Figure 3 represents the graph structure after some of the members have resolved the detected similarities and duplication.

After the resolve, it can be noticed that one of the private contacts of the most left member has been deleted because it was a real duplication in the phonebook and the owner member had found it relevant. The other one on the right side still remained because that member has not decided about it yet.

Figure 3 shows an example where four from the five similarities have been resolved (members found them relevant) and there is still one in the system (the member has not checked it yet). Resolving a similarity means that a customized link edge is being formed between the private contact(s) in one's phonebook and the relevant member who represents the same person in the system.

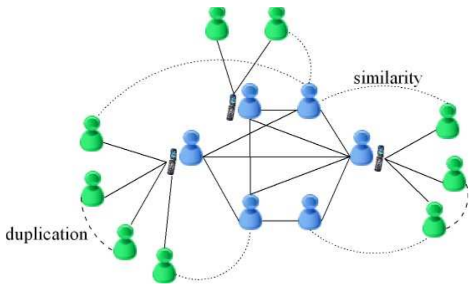

Figure 3. Resolving similarities and duplications

The private contacts that are linked to members via this type of customized links are called customized contacts. One of the key advantages of mobile based social networks are these customized links, because if a member changes his personal detail on the web user interface (adds a new phone number, uploads a new image, changes the website address, etc), it will be automatically propagated to those phonebooks, where there is a customized contact related to this member. Additional important advantages of mobile based social networks are:

- Private contacts can be managed (list, view, edit, call, etc.) from a browser.
- Similarity detecting algorithm notifies the user if duplicate contacts have been detected in the phonebook and warns about it.
- Private contacts are safely backed up in case the phone gets lost.
- Private contacts can be easily transferred to a new phone if the user replaces the old one.
- Phonebooks can be shared between multiple phones, if one happens to use more than one phone.
- It is not necessary to explicitly search for the friends in the system, because it notices if there are members similar to the private contacts in the phonebooks and notifies about it.

IV. Similarity Detecting from the Perspective of Phonebook Sizes

We took part in the implementation of a mobile based social network, called Phonebookmark at Nokia Siemens Networks which helped us to understand the dynamic behavior of the network.

Phonebookmark uses a semi-automatic similarity resolving mechanism. First it detects similarities and calculates a weight for them, which indicates how likely the corresponding phonebook contact and the member of the network identify the same person. The details of the exact comparison algorithm are discussed in [13].

The similarity detecting algorithm runs in the background on demand. If any operation occurs, which could lead to similarities; it should process them and compare the relevant entities as fast as possible. This behavior is similar to that of a queuing system where the processing unit is the algorithm and the entities in the queue are the person pairs which are waiting for comparison.

In the following model we consider the registration operation, since it can cause the most similarities when the new member uploads her or his phonebook. This operation can be divided for two main tasks. Firstly, when a member registers, she or he should be compared to every private contact in the network. If we consider the number of private contacts in a phonebook of a member as a random variable X_{Pc}, this means $E[X_{Pc}]*N_M$ comparisons, where $E[X_{Pc}]$ is the expected value of the phonebook sizes and N_M is the number of members in the network before the registration. After a reasonable operational period and initial state, the number of members N_M can be treated as a constant value.

Based on the database of Phonebookmark we were able to estimate the distribution of phonebook sizes. Figure 4 shows the tail distribution of the phonebook sizes, where the x-axis has linear scale and the *y*-axis logarithmic scale. The points on this figure fit very well to a line, which means that the tail of the distribution of the phonebook sizes decreases exponentially. This provides a simple empirical test for whether a random variable has an exponential distribution. The gradient of the function gives the parameter of the exponential distribution (Figure 4).

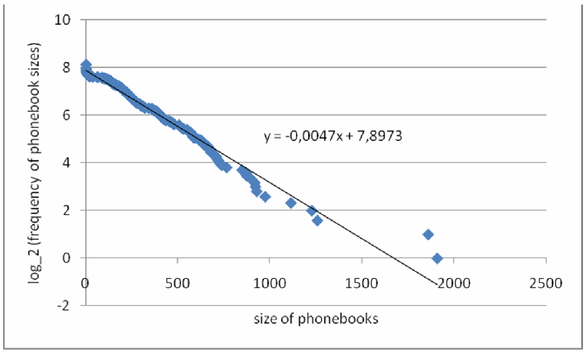

Figure 4. Size of phonebooks in Phonebookmark

In this measurement this parameter is *0.0047*, the expected value of the exponential distribution can be calculated as the reciprocal of this parameter. This way the expected value of phonebook sizes according to this measurement is *212*. Following we refer to $E[X_{Pc}]$ as C.

The other task during the member registration is to check, which members of the network are in the phonebook of the new member. This task requires N_M*X_{Pc} comparisons, where the size of the new phonebook is modeled by a random variable X_{Pc} of exponential distribution.

In this way the number of comparisons required by a member registration is modeled with the X^*_{Pc} random variable:

$$X^*_{Pc} = C * N_M + X_{Pc} * N_M = N_M * (C + X_{Pc}). \quad (1)$$

In the following we show that X*Pc has exponential distribution.

Lemma 1. The random variable X^*_{Pc} has an exponential distribution.

Proof: Because of the linear transformation, the distribution function of X^*_{Pc} looks as follows:

$$F_{X^*_{Pc}}(x^*) = F_{X_{Pc}}\left(\frac{x^* - N_M * C}{N_M}\right), \quad (2)$$

when $N_M > 0$, which is always true in our case. Hence X_{Pc} and X^*Pc have the same form, X^*_{Pc} has also exponential distribution. □

V. ESTIMATING REGISTRATION DELAY

In this Section we propose a model how to estimate the delay of similarity detection caused by member registrations. The proposed model is based on Kleinrock's queuing theory [14] which shows its relevancy in the new type of system as well.

We model the registration rate of members as a Poisson process with λ parameter. We assume that a person pair comparison takes a time unit.

Proposition. In order to keep the stability of the similarity detecting the following is required for the rate of member arrival:

$$\lambda < \frac{1}{2CN_M}. \quad (3)$$

Proof: According to Kleinrock's model for queuing systems (Section 3.2 in [14]), when the arrival rate is modeled with a λ parameter Poisson distribution and the processing rate with exponential distribution with μ parameter then the following condition is required for stability:

$$\frac{\lambda}{\mu} < 1. \quad (4)$$

This means that the expected value of serving time ($1/\mu$) must be smaller than the expected value of the time between arrivals ($1/\lambda$). In our case the expected value of the serving time is $E[X^*_{Pc}]$, since we considered a person pair comparison as the time unit. By applying Lemma 1 we can see that the serving time X^*_{Pc} has an exponential distribution and the expected value of it is calculated by:

$$E[X^*_{Pc}] = E[CN_M + X_{Pc}N_M] = CN_M + N_M E[X_{Pc}] = 2CN_M. \quad (5)$$

In case of exponential distributions, the reciprocal of the expected value is the μ parameter of the distribution. This way the requirement of the stability looks as follows:

$$\frac{\lambda}{\frac{1}{2CN_M}} < 1, \quad \lambda < \frac{1}{2CN_M}. \quad (6)$$

□

By applying the proposition, the average number of person pairs waiting for comparison Q can be calculated (Section 3.2 in [14]) as:

$$Q = \frac{\lambda}{\frac{1}{2CN_M} - \lambda}. \quad (7)$$

Based on this model, the resource requirement of the similarity detecting canbe calculated in real environment, considering the speed of the processing unit(s). In order to demonstrate the behavior of this queue, we have made measurements regarding to the registration of the members in Phonebookmark. Figure 5 illustrates the queue length considering $2CN_M$ and $2.5CN_M$ processing rates.

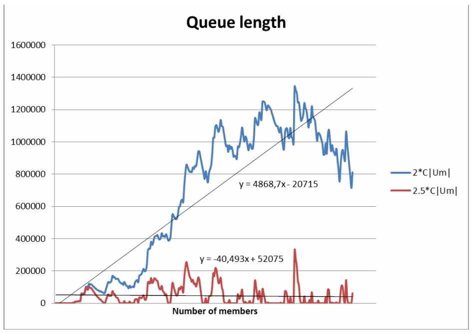

Figure 5. Queue length for similarity detecting

The *x*-axis shows as the number of members in the system increases, while the *y*-axis represents the number of comparison steps when a new member registers (sum of the remaining comparison and the new ones). It can be seen that the queue length can be decreased significantly, when the processing rate increases.

Figure 6 illustrates the queue length normalized with the number of members.

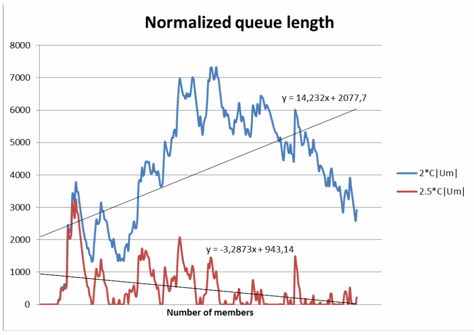

Figure 6. Normalized queue length for similarity detecting

VI. CONCLUSION

Social network sites are becoming more and more important in everyday life. These networks enable to manage relationships online. From another perspective they are an environment created by the people who are using them. The advances in web technologies also aver these solutions. Besides that the increasing hardware and software capabilities of mobile devices enable them to connect to these networks and there are already huge amount of clients for mobile phones related to social networking sites. These solutions consider mobile phones mainly as simple clients, or maybe they allow some special functions for them like instant image upload.

If we consider the fact that the phonebook of the mobile phones represents our real life relationships, then mobile phones get a special role in social networking solutions. In this paper we briefly summarized mobile related social networks. They allow synchronizing the phonebook to the social network and during this process they discover similarities automatically between members of the network and the phonebook entries. In this way the system can automatically find our friends in the social network. Besides that such social networks have a more dynamical behavior, they can keep the phonebooks always up to date with updated contact information provided by our friends. The solution considers privacy issues as well, as users can determine the visibility of their personal details and private contacts.

As a main contribution of this paper we examined the distribution of phonebook sizes and we proposed a model for registration delay considering similarity detecting, which helps to design the resource requirements of such systems. The future plan includes analyzing the dynamic behavior and growth of such system and examining how to integrate mobile communications also in the network.

ACKNOWLEDGMENT

This work is supported by the New Hungary Development Plan (Project ID: TÁMOP-4.2.1/B-09/1/KMR-2010-0002 and TAMOP 4.2.1/B-09/1/KMR-2010-0003).

REFERENCES

[1] Alexa. http://www.alexa.com/topsites. June 2010.

[2] Facebook statistics, http://www.facebook.com/press/info.php?statistics, February, 2010.

[3] D. M. Boyd and N. B. Ellison: Social network sites: Definition, history, and scholarship, Journal of Computer-Mediated Communication, Volume 13, Issue 1 (2007)

[4] M. E. J. Newman, D. J. Watts, and S. H. Strogatz: Random graph models of social networks. Proc. National Academy of Sciences, 2002.

[5] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee: Measurement and analysis of online social networks. In: ACM/USENIX IMC, 2007.

[6] A. Broder, R. Kumar, F. Maghoul, P. Raghavan, S. Rajagopalan, R. Stata, A. Tomkins, and J. Wiener: Graph structure in the web. In: Proc. of the 9th international World Wide Web conference on Computer networks, 2000.

[7] N. S. Raza and C.-N. Chuah: Unveiling Facebook: A measurement Study of Social Network Based Applications. In: Proc. ACM Internet Measurement Conference (IMC), 2008.

[8] R. Kumar and J. Novak, A. Tomkins: Structure and evolution of online social networks. In: KDD'06, 2006.

[9] Zyb social networking application. http://www.zyb.com, 12th of June, 2010.

[10] Plaxo social networking application. http://www.plaxo.com, 13th of June 2010.

[11] Xing social networking application. http://www.xing.com, 15th of June 2010.

[12] P. Ekler and T. Lukovszki. Similarity Distribution in Phonebook-centric Social Networks. In: 5th International Conference on Wireless and Mobile Communications (ICWMC). 2009.

[13] P. Ekler and T. Lukovszki: Learning Methods for Similarity Handling in Phonebook-centric Social Networks. In: 10th International Symposium of Hungarian Researchers on Computational Intelligence and Informatics (CINTI 2009), 2009.

[14] L. Kleinrock: Queueing Systems. Volume 1: Theory, Wiley-Interscience, pages 94-101, 1975.